BOEING SCIENTIFIC RESEARCH LABORATORIES DOCUMENT D1-82-0578

BOEING

SCIENTIFIC RESEARCH LABORATORIES

A Multivariate Notion of Association for General Random Variables

J. D. Esary

Frank Proschan

D. W. Walkup

also available from Author

MATHEMATICS RESEARCH OCTOBER 1966

D1-82-0578

A MULTIVARIATE NOTION OF ASSOCIATION FOR GENERAL RANDOM VARIABLES

by

J. D. Esary

Frank Proschan

D. W. Walkup

Mathematical Note No. 490

Mathematics Research Laboratory

BOEING SCIENTIFIC RESEARCH LABORATORIES

October 1966

Abstract

In a previous paper [A multivariate notion of association, with a reliability application] random variables $T_1, T_2, \ldots, T_n$ were defined to be *associated* if each pair of non-decreasing functions $F(T_1, T_2, \ldots, T_n)$, $G(T_1, T_2, \ldots, T_n)$ have a non-negative covariance. The properties of this definition were studied in the case that $T_1, T_2, \ldots, T_n$ are finitely discrete, and a sample application to reliability theory was discussed. In the present paper several equivalent definitions of the same notion of association for unrestricted $T_1, T_2, \ldots, T_n$ are treated. The properties previously obtained, i.e. that association is preserved under the operations of extracting subsets, pooling independent sets, and forming sets of non-decreasing functions, are shown to hold in general. In addition, association is shown to be preserved under limits in distribution. Some additional applications of association are discussed, e.g. previously published results of A. W. Kimball and H. Robbins are obtained.

Acknowledgments

We are grateful to V. Klee, A. W. Marshall, and R. Pyke for a number of helpful suggestions.

BLANK PAGE

1. Introduction

In a previous paper [1] we studied a notion of multivariate association for random variables using the definition

(1.1) Random variables $T_1, T_2, \ldots, T_n$ are *associated* if $\mathrm{Cov}[F(\underset{\sim}{T}), G(\underset{\sim}{T})] \geq 0$, where $\underset{\sim}{T} = (T_1, T_2, \ldots, T_n)$, for all pairs F, G of non-decreasing functions,

where F non-decreasing means $F(\underset{\sim}{s}) \leq F(\underset{\sim}{t})$ whenever $s_i \leq t_i$, $i=1,\ldots,n$. We considered only the case in which the random variables $T_1, T_2, \ldots, T_n$ are finitely discrete, i.e. take a finite number of values. For such variables the expectations $EF(\underset{\sim}{T})$, $EG(\underset{\sim}{T})$, and $EF(\underset{\sim}{T})G(\underset{\sim}{T})$ involved in $\mathrm{Cov}[F(\underset{\sim}{T}), G(\underset{\sim}{T})]$ always exist, and (1.1) is consequently a well-behaved definition.

We showed in the finitely discrete case that association has a number of fundamental properties:

(P_1) *Any subset of a set of associated random variables is also a set of associated random variables.*

(P_2) *If two sets of associated random variables are independent of one another, then their union is a set of associated random variables.*

(P_3) *The set consisting of a single random variable is associated.*

(P_4) *If* $T_1, T_2, \ldots, T_n$ *are associated, then any set of non-decreasing functions* $S_1(\underset{\sim}{T}), S_2(\underset{\sim}{T}), \ldots, S_m(\underset{\sim}{T})$ *are associated.*

Some discussion of the implications of the properties of association, and a sample application, are given in [1]. The purpose of the present paper is to consider several "well-behaved" definitions of association for general, i.e. unrestricted, random variables, and to show that they are equivalent; to show that properties P_1, P_2, P_3, and P_4 hold in the general case, and to obtain the additional property

(P_5) *If* $T_1^{(k)}, T_2^{(k)}, \ldots, T_n^{(k)}$ *are associated for each* k *and* $\underset{\sim}{T}^{(k)} \to \underset{\sim}{S} = (S_1, S_2, \ldots, S_n)$ *in distribution, then* $S_1, S_2, \ldots, S_n$ *are associated;*

and finally to indicate that a suitably qualified version of (1.1) could be used as a definition of association in the general case.

2. Equivalent definitions of association for general random variables

Perhaps the most elementary way to define association for unrestricted random variables is

(2.1) Random variables $T_1, T_2, \ldots, T_n$ are *associated* if $\mathrm{Cov}[\Gamma(\underset{\sim}{T}), \Delta(\underset{\sim}{T})] \geq 0$ for all pairs Γ, Δ of binary, non-decreasing functions.

In [1] definition (2.1) is shown equivalent to definition (1.1) for finitely discrete random variables.

We will also consider the definition

(2.2) Random variables $T_1, T_2, \ldots, T_n$ are *associated* if $\mathrm{Cov}[U(\underset{\sim}{T}), V(\underset{\sim}{T})] \geq 0$ for all pairs U, V of bounded, continuous, non-decreasing functions.

Finally we introduce, for each random variable T_i and fixed t, the binary indicator variable $X_i(t)$ such that $X_i(t) = 1$ if $T_i > t$, $X_i(t) = 0$ if $T_i \leq t$. We then consider the definition

(2.3) Random variables $T_1, T_2, \ldots, T_n$ are *associated* if, for every choice of k and $t_1, t_2, \ldots, t_k$, the binary random variables in the array

$$\begin{matrix} X_1(t_1) & X_1(t_2) & \cdots & X_1(t_k) \\ X_2(t_1) & X_2(t_2) & \cdots & X_2(t_k) \\ \vdots & & & \vdots \\ X_n(t_1) & X_n(t_2) & \cdots & X_n(t_k) \end{matrix}$$

are associated according to (2.1).

Property P_1 is an immediate consequence of each of the definitions (2.1), (2.2), and (2.3). The proof of property P_2 given in [1] can be applied to show it holds for each of (2.1), (2.2), and (2.3). In [1] property P_3 was proved for definition (2.3). The proof of property P_4 given in [1] applies to show that it holds for definition (2.1). Property P_5 holds for definition (2.2), by application of the Helly-Bray theorem.

We will show that *the three definitions* (2.1), (2.2), *and* (2.3) *are equivalent.* Our plan for the proof will be to show that association according to (2.3) implies association according to (2.2), and then to show that association according to (2.2) implies association according to (2.1). That association by (2.1) implies association by (2.3) is immediate, since property P_4 holds for (2.1), and each binary indicator $X_i(t)$ is a non-decreasing function of T_i. Given that definitions (2.1), (2.2), and (2.3) are equivalent, it follows that *properties* P_1, P_2, P_3, P_4, *and* P_5 *hold for each of the definitions,* since each property has been shown to hold for at least one definition.

Theorem 2.1. *If* $T_1,T_2,\ldots,T_n$ *are associated according to definition* (2.3), *then* $\mathrm{Cov}[U(\underset{\sim}{T}),V(\underset{\sim}{T})] \geq 0$ *for all pairs* U,V *of bounded, continuous, non-decreasing functions.*

Proof. Since we can add large enough constants to each of a pair of bounded functions to make them non-negative without changing their covariance, it suffices to prove the theorem for pairs U,V of non-negative, bounded, continuous, non-decreasing functions.

Since the binary random variables in the array $\underset{\sim}{X}(\underset{\sim}{t}) = \{X_i(t_j);\ i=1,\ldots,n, j=1,\ldots,k\}$ are associated according to (2.1), $\mathrm{Cov}[F\underset{\sim}{X}(\underset{\sim}{t}),G\underset{\sim}{X}(\underset{\sim}{t})] \geq 0$, for any pair F,G of non-decreasing functions (by our previous observation that definition (2.1) implies definition (1.1) in the finitely discrete case).

Consider some choice of $t_1 < t_2 < \cdots < t_k$, and define, e.g. for U, $F^{(k)}(\underset{\sim}{T}) = 0$ if any $X_i(t_1) = 0$, $i=1,\ldots,n$; $F^{(k)}(\underset{\sim}{T}) = U(\underset{\sim}{S})$, where $S_i = \max\{t_j | X_i(t_j)=1\}$, if all $X_i(t_1) = 1$. The functions $F^{(k)}$ are non-negative and *non-decreasing,* viewed either as functions of $\underset{\sim}{T}$ or as functions of $\underset{\sim}{X}(\underset{\sim}{t})$. Now let $\{t_1,t_2,\ldots,t_k\}$ increase with k to a countable, dense set in $(-\infty,+\infty)$. Since U,V, and $U \cdot V$ are continuous, $0 \leq F^{(k)} \uparrow U$, $0 \leq G^{(k)} \uparrow V$, and $0 \leq F^{(k)} \cdot G^{(k)} \uparrow U \cdot V$ at each fixed value of $\underset{\sim}{T}$. Since U,V, and $U \cdot V$ are bounded, by monotone convergence $EF^{(k)}(\underset{\sim}{T}) \uparrow EU(\underset{\sim}{T})$, $EG^{(k)}(\underset{\sim}{T}) \uparrow EV(\underset{\sim}{T})$, and $EF^{(k)}(\underset{\sim}{T}) \cdot G^{(k)}(\underset{\sim}{T}) \uparrow EU(\underset{\sim}{T}) \cdot V(\underset{\sim}{T})$. Thus $0 \leq \mathrm{Cov}[F^{(k)}(\underset{\sim}{T}),G^{(k)}(\underset{\sim}{T})] \to \mathrm{Cov}[U(\underset{\sim}{T}),V(\underset{\sim}{T})]$.

__Lemma__ 2.2. If $T_1,T_2,\ldots,T_n$ are associated according to definition (2.2), then $\mathrm{Cov}[\Phi(\underset{\sim}{T}),\Psi(\underset{\sim}{T})] \geq 0$ for all pairs Φ,Ψ of binary, right continuous, non-decreasing functions.

__Proof__. Consider e.g. Φ. Let $A = \{\underset{\sim}{t}|\Phi(\underset{\sim}{t})=1\}$. Let $d(\underset{\sim}{t},A)$ be the Euclidean distance from a point $\underset{\sim}{t}$ to the set A. Define $U^{(k)}(\underset{\sim}{T}) = 0$ if $d(\underset{\sim}{T},A) \geq 1/k$, $U^{(k)}(\underset{\sim}{T}) = 1 - k \cdot d(\underset{\sim}{T},A)$ if $d(\underset{\sim}{T},A) \leq 1/k$. Each function $U^{(k)}$ is non-negative, *bounded, continuous,* and *non-decreasing*. Since Φ is right continuous, A is closed. We have $U^{(k)}(\underset{\sim}{t}) = 1$, if $\underset{\sim}{t} \in A$, all k, and $U^{(k)}(\underset{\sim}{t}) \searrow 0$, as $k \to \infty$, if $\underset{\sim}{t} \in A^c$. Thus $1 \geq U^{(k)} \searrow \Phi$, as $k \to \infty$.

As in Theorem 2.1, by monotone convergence, we conclude that $\mathrm{Cov}[\Phi(\underset{\sim}{T}),\Psi(\underset{\sim}{T})] \geq 0$.

__Theorem__ 2.3. *If* $T_1,T_2,\ldots,T_n$ *are associated according to definition* (2.2), *then* $T_1,T_2,\ldots,T_n$ *are associated according to definition* (2.1).

__Proof__. Consider e.g. Γ. Let $A = \{\underset{\sim}{t}|\Gamma(\underset{\sim}{t})=1\}$. We can find a compact set $C \subset A$ such that $P[C] + \varepsilon \geq P[A]$. Let $C^+ = \{\underset{\sim}{c} + \underset{\sim}{t}|\underset{\sim}{c} \in C, t_1 \geq 0,\ldots,t_n \geq 0\}$. Then $C \subset C^+ \subset A$ and C^+ is closed.

Let $\Phi(\underset{\sim}{T}) = 1$ if $\underset{\sim}{T} \varepsilon C^{+}$, $\Phi(\underset{\sim}{T}) = 0$ if $\underset{\sim}{T} \varepsilon (C^{+})^{c}$. The function Φ is *binary, right continuous,* and *non-decreasing;* $\Gamma \geq \Phi$; and $E\Phi(\underset{\sim}{T}) + \varepsilon \geq E\Gamma(\underset{\sim}{T})$. Thus $\mathrm{Cov}[\Gamma(\underset{\sim}{T}),\Delta(\underset{\sim}{T})]$

$\geq E\Phi(\underset{\sim}{T})\Psi(\underset{\sim}{T}) - \{E\Phi(\underset{\sim}{T}) + \varepsilon\}\cdot\{E\Psi(\underset{\sim}{T}) + \varepsilon\} \geq \mathrm{Cov}[\Phi(\underset{\sim}{T}),\Psi(\underset{\sim}{T})] - 2\varepsilon - \varepsilon^2$.

Letting $\varepsilon \to 0$ and using Lemma 2.2, we obtain $\mathrm{Cov}[\Gamma(\underset{\sim}{T}),\Delta(\underset{\sim}{T})] \geq 0$.

3. The covariance of arbitrary, non-decreasing functions

We now prove a theorem which indicates that we could define association for general random variables in essentially the same way, i.e. by (1.1), that we defined it for finitely discrete random variables.

Theorem 3.1. *If* $T_1,T_2,\ldots,T_n$ *are associated and* F,G *are a pair of non-decreasing functions such that* $EF(\underset{\sim}{T})$, $EG(\underset{\sim}{T})$, *and* $EF(\underset{\sim}{T})G(\underset{\sim}{T})$ *exist, then* $\mathrm{Cov}[F(\underset{\sim}{T}),G(\underset{\sim}{T})] \geq 0$.

Proof. That $T_1,T_2,\ldots,T_n$ are associated implies, by P_4, that $F(\underset{\sim}{T})$, $G(\underset{\sim}{T})$ are associated, which in turn implies, by P_4, that $X_F(s)$, $X_G(t)$ are associated for all s, t, where e.g. $X_F(s) = 1$ if $F(\underset{\sim}{T}) > s$, $X_F(s) = 0$ if $F(\underset{\sim}{T}) \leq s$. Thus $\mathrm{Cov}[X_F(s),X_G(t)] \geq 0$, all s, t. The result follows from the identity

$$\mathrm{Cov}[F(\underset{\sim}{T}),G(\underset{\sim}{T})] = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} \mathrm{Cov}[X_F(s),X_G(t)]\,ds\,dt. \tag{3.1}$$

[See Lehmann (1966), who gives a convenient statement of (3.1) and a proof attributed to W. Hoeffding. See also Marshall-Olkin (1966) for a comparable result on moment generating functions.]

4. Applications

Several interesting applications may be obtained as a consequence of

Theorem 4.1. *Let* $T_1, T_2, \ldots, T_n$ *be associated. Then*

$$(4.1a)\quad P[T_1 \le t_1, T_2 \le t_2, \ldots, T_n \le t_n] \ge \prod_{i=1}^{n} P[T_i \le t_i]$$

and

$$(4.1b)\quad P[T_1 > t_1, T_2 > t_2, \ldots, T_n > t_n] \ge \prod_{i=1}^{n} P[T_i > t_i],$$

for all $t_1, t_2, \ldots, t_n$.

Proof. Let $X_i(t_i) = 1$ if $T_i > t_i$. Then the binary random variables $X_1(t_1), X_2(t_2), \ldots, X_n(t_n)$ are associated by P_4. Using (2.3.2) and (2.3.1) of [1] we obtain (4.1a) and (4.1b), respectively.

Partial sums [Robbins (1954)]

Let $T_1, T_2, \ldots, T_n$ be independent random variables. Then $T_1, T_2, \ldots, T_n$ are associated by P_2 and P_3. Let $S_j = \sum_{i=1}^{j} T_i$, $j=1,\ldots,n$. Then $S_1, S_2, \ldots, S_n$ are associated by P_4. Thus

$$P[S_1 \le s_1, S_2 \le s_2, \ldots, S_n \le s_n] \ge \prod_{j=1}^{n} P[S_j \le s_j]$$

for all $s_1, s_2, \ldots, s_n$ by (4.1a).

Order statistics

Let $T_1, T_2, \ldots, T_n$ be a sample of size n, and $S_1, S_2, \ldots, S_n$ be the corresponding order statistics. Then $T_1, T_2, \ldots, T_n$ are independent and thus associated by P_2 and P_3. Since each S_j is a non-decreasing function of $T_1, T_2, \ldots, T_n$, $S_1, S_2, \ldots, S_n$ are associated by P_4. Thus (4.1a) and (4.1b) hold for $S_1, S_2, \ldots, S_n$, and by P_1 also for every subset of $S_1, S_2, \ldots, S_n$.

Multivariate exponential distribution

Marshall and Olkin (1966) consider the multivariate exponential distribution with the distribution function $F(s_1, s_2, \ldots, s_m)$ such that

$$(4.2)\quad 1 - F(s_1, s_2, \ldots, s_m) = \exp[-\sum_{j=1}^{m} \lambda_j s_j - \sum_{j<k} \lambda_{jk} \max(s_j, s_k) - \sum_{j<k<\ell} \lambda_{jk\ell} \max(s_j, s_k, s_\ell) - \cdots - \lambda_{12\ldots m} \max(s_1, s_2, \ldots, s_m).$$

As pointed out by Marshall and Olkin if random variables $S_1, S_2, \ldots, S_m$ are distributed according to (4.2), then there exist independent, exponentially distributed random variables $T_1, T_2, \ldots, T_n$ such that $S_j = \min(T_i; i \in A_j)$, where $A_j \subset \{1, 2, \ldots, n\}$. Since $T_1, T_2, \ldots, T_n$ are independent and each S_j is a non-decreasing function of $T_1, T_2, \ldots, T_n$, we obtain from Theorem 4.1

$$(4.3a) \qquad F(s_1, s_2, \ldots, s_m) \geq \prod_{j=1}^{m} F_j(s_j)$$

and

$$(4.3b) \qquad 1 - F(s_1, s_2, \ldots, s_m) \geq \prod_{j=1}^{m} [1-F_j(s_j)],$$

where F_j is the marginal distribution function of S_j. Marshall and Olkin discuss inequalities (4.3a) and (4.3b) for the bivariate exponential, and give a further, quantitative analysis of that case.

<u>Analysis of variance</u>

Kimball (1951) considers the case of analysis of variance in which two hypotheses are tested using the same error variance for each test. As an example of particular importance, he cites the case in which the effects of both rows and columns are to be tested. As usually formulated, three quadratic forms, q_1, q_2, q_3, are computed, independently distributed as χ^2 with n_1, n_2, n_3 degrees of freedom respectively, q_1 representing the sum of squares between rows, q_2 the sum of squares between columns, and q_3 the sum of squares due to error. The likelihood ratio test statistics for testing the two hypotheses are

$$F_1 = \frac{q_1/n_1}{q_3/n_3} \quad \text{and} \quad F_2 = \frac{q_2/n_2}{2_3/n_3} .$$

The probability of making no errors of the first kind is $P[F_1 < F_{1\alpha}, F_2 < F_{2\alpha}]$, where $F_{1\alpha}(F_{2\alpha})$ is the 100α per cent point of the distribution of $F_1(F_2)$. Kimball proves

$$(4.4) \qquad P[F_1 \leq F_{1\alpha}, F_2 \leq F_{2\alpha}] \geq P[F_1 \leq F_{1\alpha}]P[F_2 \leq F_{2\alpha}].$$

In other words, the assurance of no errors of the first kind is greater following the standard experimental procedure than if separate experiments had been performed.

Kimball's result is an immediate consequence of Theorem 4.1 if we note that q_1, q_2, q_3^{-1} are associated (since independent), and F_1, F_2 are non-decreasing functions of q_1, q_2, q_3^{-1}.

References

[1] Esary, J. D., Proschan, F., and Walkup, D. W. (1966). A multivariate notion of association, with a reliability application. Boeing Scientific Research Laboratories Document D1-82-0567.

[2] Kimball, A. W. (1951). On dependent tests of significance in the analysis of variance. *Ann. Math. Statist.* *22*, 600-602.

[3] Lehmann, E. L. (1966). Some concepts of dependence. *Ann. Math. Statist.* *37*, 1137-1153.

[4] Marshall, A. W. and Olkin, I. (1966). A multivariate exponential distribution. Boeing Scientific Research Laboratories Document D1-82-0505; to be published in *J. Amer. Statist. Assoc.*

[5] Robbins, H. (1954). A remark on the joint distribution of cumulative sums. *Ann. Math. Statist.* *25*, 614-616.